

LearnOSM

Gegevens van OSM nakijken

Bijgewerkt op 24-04-2017

Table of Contents

Gegevens van OSM nakijken.....	1
Dagelijkse controles.....	2
Valideren van gegevens.....	3
JOSM Zoeken gebruiken.....	6
Opnieuw onderzoeken.....	9
Fouten corrigeren.....	9
Rapporteren over nauwkeurigheid.....	9
SQL-query's.....	10
De gegevens voorbereiden.....	10
SQL-query's.....	14
Samenvatting.....	17

Dit gedeelte behandelt processen voor het controleren van de kwaliteit van gegevens, in het bijzonder in de context van een gecoördineerd project voor in kaart brengen in OSM, zoals die welke worden ondernomen door het [Humanitarian OpenStreetMap Team](#) in verscheidene landen en projecten van [Open Cities](#) in Bangladesh, Sri Lanka, en Nepal. De gedemonstreerde methoden kunnen ook nuttig zijn in een andere contexten, waar het nakijken van de kwaliteit van gegevens een normale taak is.

Wanneer we proberen een volledige set van objecten en attributen in een gespecificeerd gebied in kaart te brengen, hebben we manieren nodig om te controleren op fouten en manieren om toegang te krijgen tot de nauwkeurigheid van het werk. In deze handleiding zullen we verschillende methoden voor het controleren van gegevens behandelen, de stappen van de methoden uitleggen en de reden daarvoor. Een goed beheerd project voor in kaart brengen zal al deze drie processen bevatten, zowel voor het evalueren en corrigeren van gegevens als voor rapporteren.

- Dagelijkse controles
- Opnieuw onderzoeken
- SQL-query's

Deze methoden voor nakijken worden belangrijker naarmate het gegevensmodel groeit en het aantal verzamelde objecten behoorlijk groot begint te worden. Het zou bijvoorbeeld heel veel tijd en inspanning vergen om een gegevensmodel te beoordelen dat alleen points of interest (POI's) betreft:

Data Model	
POIs	Name
	Type
	Phone Number

In dit geval zouden de te stellen vragen moeten zijn:

- Zijn de POI's op alle locaties in kaart gebracht?
- Zijn er POI's waar het attribuut name ontbreekt?
- Zijn er POI's waar het attribuut type ontbreekt?
- Zijn er POI's waar het attribuut phone number ontbreekt?
- Is de waarde in het veld name op de juiste wijze voorzien van hoofdletters?
- Is het telefoonnummer een geldig nummer?

Gewoonlijk is het gegevensmodel echter veel meer complex, zoals in het geval met het in kaart brengen van gebouwen. Denk aan een gegevensmodel dat dit bevat:

Buildings	Name
	Street Name
	Building Number
	Number of Levels
	Use of Building
	Material of Building
	Type of Structure
	Use of Building
	Construction Date
	Building Condition

Nu zou u duizenden gebouwen in kaart brengen die vele attributen hebben, en de analyse wordt kritischer. In deze handleiding zullen we gebouwen als voorbeeld gebruiken, hoewel dezelfde methoden ook kunnen worden toegepast voor het nakijken van andere typen objecten.

Dagelijkse controles

De meest directe weg om gegevens te controleren is door ze op met een bepaalde regelmaat na te kijken en te valideren. Dit zou dagelijks kunnen zijn of op zijn hoogst per week. Voor de supervisor van een team van mensen die in kaart brengen is dit een belangrijke taak omdat het in een vroeg stadium ondervangen van fouten en manieren van slecht verwerken betekent dat zij kunnen worden gecorrigeerd en de verwerkers kunnen leren om de dingen op de juiste manier te doen.

Hier zullen we kijken naar enkele methoden voor het controleren van gegevens door eenvoudigweg JOSM te gebruiken. Enkele van de vragen die we voor onze gegevens zullen stellen zijn:

- Zijn er fouten in de **topologie** (zoals overlappende gebouwen of relaties die niet juist zijn)?
- Zijn er fouten bij het **taggen** (fout gespelde tags, verkeerd gebruikte combinaties van

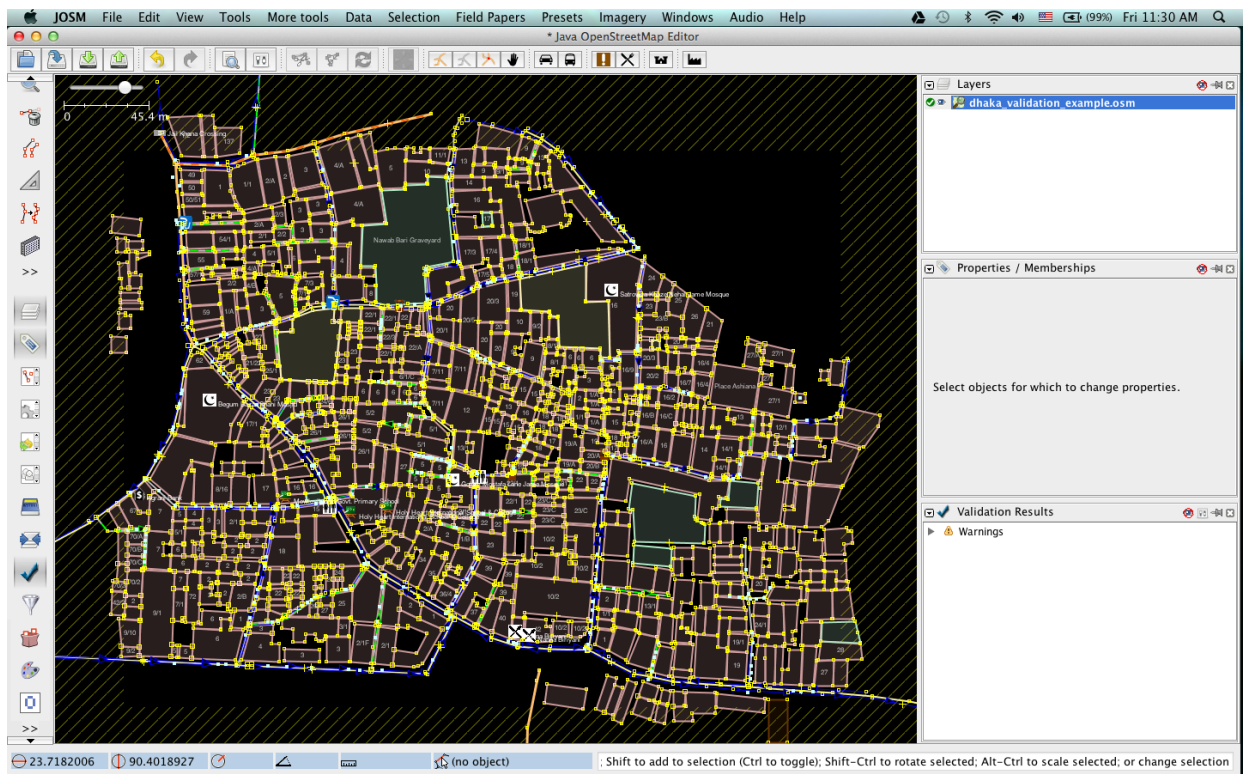
sleutel-waarde)?

- Zijn de gegevens **compleet**, overeenkomstig het gegevensmodel?

Laten we eens onderzoeken hoe we in JOSM antwoord op deze vragen kunnen krijgen. We gaan er van uit dat we het werk van anderen onderzoeken, maar dezelfde processen werken ook prima (en zouden gemakkelijker moeten zijn) als u uw eigen werk analyseert.

We zullen een voorbeeld gegevensbestand gebruiken uit het project voor het in kaart brengen van de Open Cities in Dhaka. Download, om de handleiding te kunnen volgen, het volgende bestand: dhaka_validation_example.osm

Probeer N I E T uw wijzigingen op te slaan op OpenStreetMap. Deze oefeningen zijn alleen ter demonstratie.



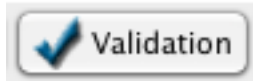
Valideren van gegevens

De eerste stap voor het controleren van gegevens is het uitvoeren van het gereedschap Validatie in JOSM, wat automatisch de gegevens, die u heeft geopend, zal controleren op verdachte fouten. Dit gereedschap is in het bijzonder geschikt voor het vinden van fouten in **topologie**, maar zou niet zo nuttig kunnen zijn voor het vinden van onjuiste tags.

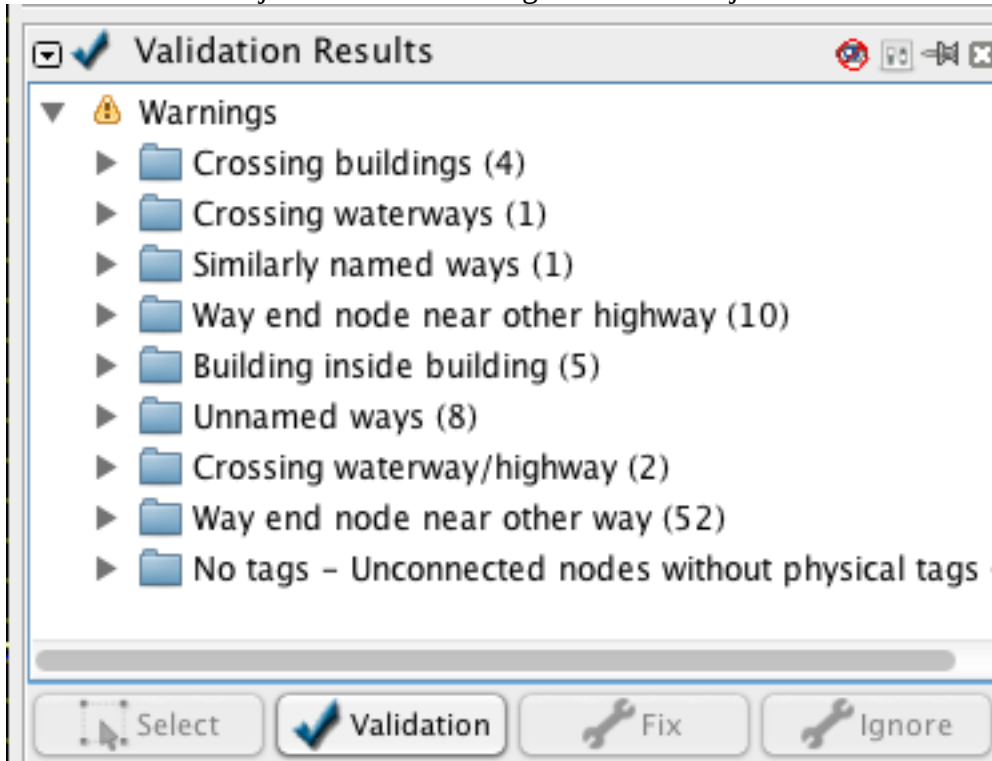
- Activeer het gereedschap door te klikken op de knop Validatie aan de linkerkant van JOSM. (Dit is niet nodig als het paneel Validatie al is geopend)



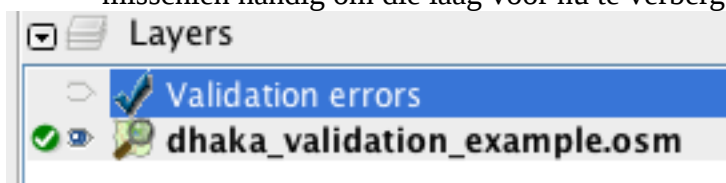
- Zorg er vervolgens voor dat er niets is geselecteerd door te klikken op een lege plek op uw kaart. Als u objecten hebt geselecteerd wanneer u het gereedschap Validatie uitvoert, zullen alleen die geselecteerde objecten worden gecontroleerd. (soms wilt u misschien alleen bepaalde objecten controleren, maar voor nu zullen we het gehele bestand gaan controleren)
- Klik op de knop “Validatie” in het paneel.



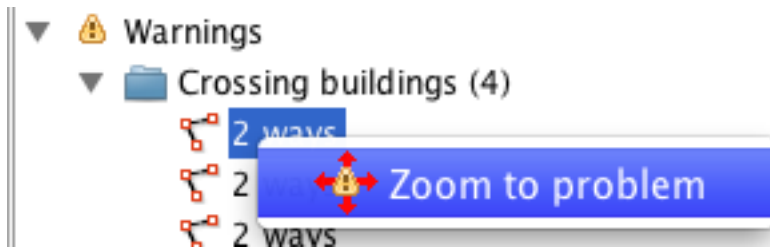
- U zult een lijst met waarschuwingen zien verschijnen:



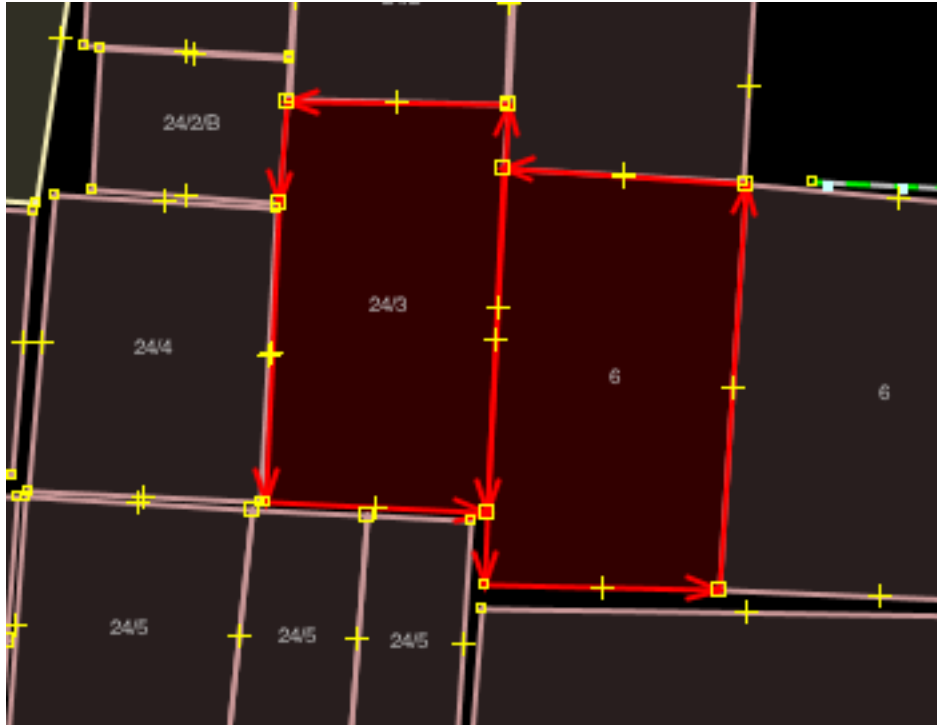
- Er verschijnt ook een nieuwe laag die weergeeft waar de fouten zijn. U vindt het misschien handig om die laag voor nu te verbergen.



Laten we eens naar een aantal waarschuwingen kijken. U ziet dat er vier waarschuwingen zijn voor “Kruist gebouwen”. Deze waarschuwing betekent dat gebouwen elkaar ergens overlappen. Selecteer het eerste item in deze lijst, klik met rechts en klik op “Inzoomen op probleem.”



Klik ook op de knop “Selecteren” aan de onderzijde van het venster Validatieresultaten om de betreffende wegen te selecteren. Dit geeft weer dat deze twee wegen een probleem hebben:

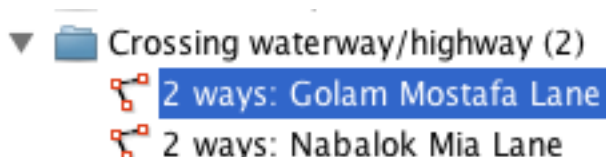


- Dit is een fout die we nooit zouden hebben afgevangen zonder het gereedschap Validatie. Als u heel ver inzoomt zult u zien dat er een kleine overlap is tussen de de gebouwen, wat een topologische fout is, omdat gebouwen elkaar gewoonlijk niet overlappen. De middelste knoop dient te worden verplaatst om dit op te lossen. Als de gebouwen elkaar in feite raken, wat zij waarschijnlijk doen, kan de middelste knoop worden samengevoegd met de weg.
- Wanneer dit is gecorrigeerd kunnen we het gereedschap Validatie opnieuw uitvoeren en deze waarschuwing zal zijn verdwenen uit de lijst.

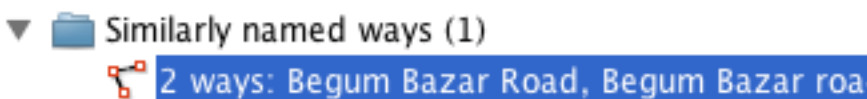
Deze methode van automatisch controleren van de gegevens is een effectieve manier voor het corrigeren van fouten in de topologie, in het bijzonder die welke door een persoon moeilijk te onderscheiden zouden zijn. In de lijst met waarschuwingen van Validatie kunt u zien dat andere waarschuwingen, zoals “Gebouw binnen gebouw” het resultaat is van een soortgelijke fout.

Toch zijn andere waarschuwingen, zoals “Kruisende waterweg/weg,” niet noodzakelijkerwijze fouten. Dit geeft aan dat het gereedschap validatie goed is in het vinden van mogelijke fouten,

maar het vereist wel dat iemand daar naartoe gaat en kijkt of de fout belangrijk is of niet.



Laten we eens kijken naar de waarschuwing onder “Soortgelijk benoemde wegen” om te zien dat een fout niet topologisch is. Klik op “Selecteren” om de twee betreffende wegen te selecteren.

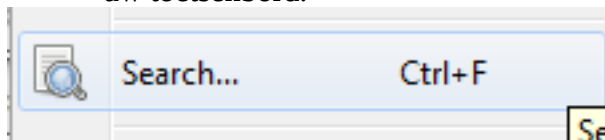


Ziet u wat de fout is? Hier hebben we twee verschillende wegsegmenten, die in feite dezelfde weg zijn, maar zij zijn enigszins afwijkend van elkaar benoemd - “road” is met een hoofdletter op één van de wegen, maar niet op de andere. Het zou logisch zijn dat zij dezelfde naam zou hebben, en in dit geval zou het woord “road” een hoofdletter moeten hebben.

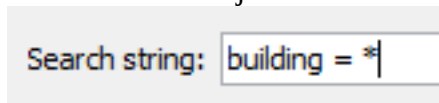
JOSM Zoeken gebruiken

Zoeken in JOSM is een krachtige manier om gegevens na te kijken. Het stelt u in staat zoektermen op te geven, ook wel bekend als query's, om alleen de objecten te selecteren die u wilt.

- Ga, om toegang te krijgen tot Zoeken, naar Bewerken -> Zoeken of druk op CTRL + F op uw toetsenbord.



- Er zijn een groot aantal typen query's waarmee u hier kunt zoeken, en u kunt details en voorbeelden zien in het zoekvak zelf en door te klikken op de knop “Help”.
- laten we, voor nu, alle gebouwen proberen te selecteren. Bijna elk gebouw zal de tag **building=yes** hebben en enkele zullen **building=construction** hebben. Dat betekent dat we een query kunnen bouwen die leest als:
building = yes OR building=construction
- Dit zou alle gebouwen moeten selecteren, maar voor het geval iemand een verkeerde tag aan een gebouw heeft toegekend, kunnen we in plaats daarvan een jokerteken gebruiken, dat alle objecten zal selecteren die de sleutel **building** hebben.

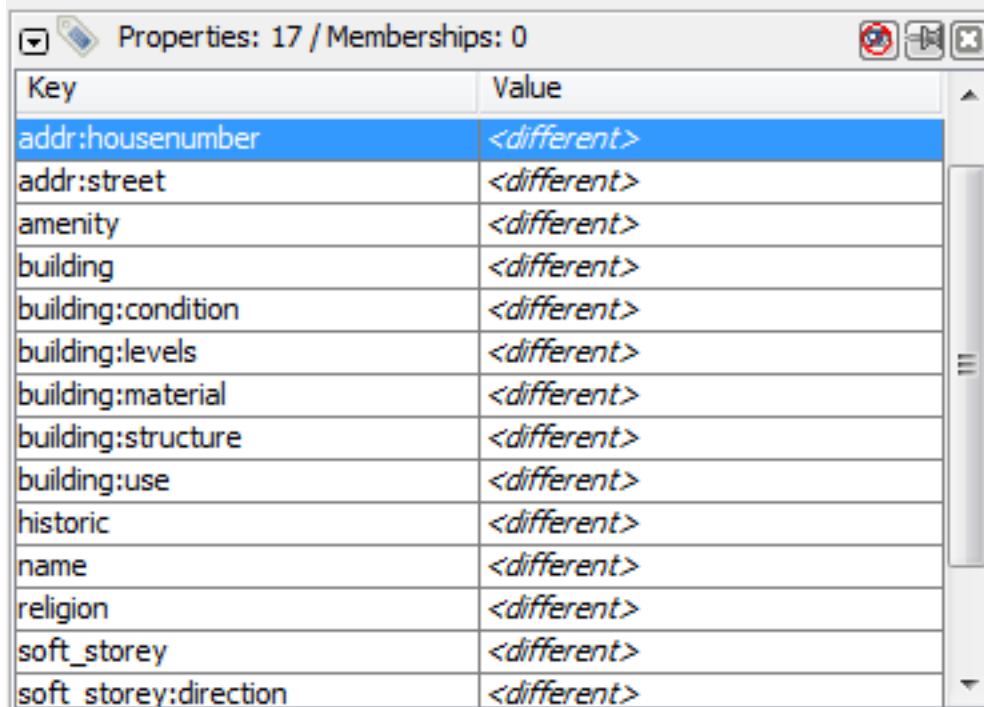


- Alle gebouwen zullen worden geselecteerd.

Dat is fantastisch, maar hoe helpt dat ons om de gegevens na te kijken? Wel, nu alle objecten van één type zijn geselecteerd, kunnen we zoeken naar niet juiste tags.

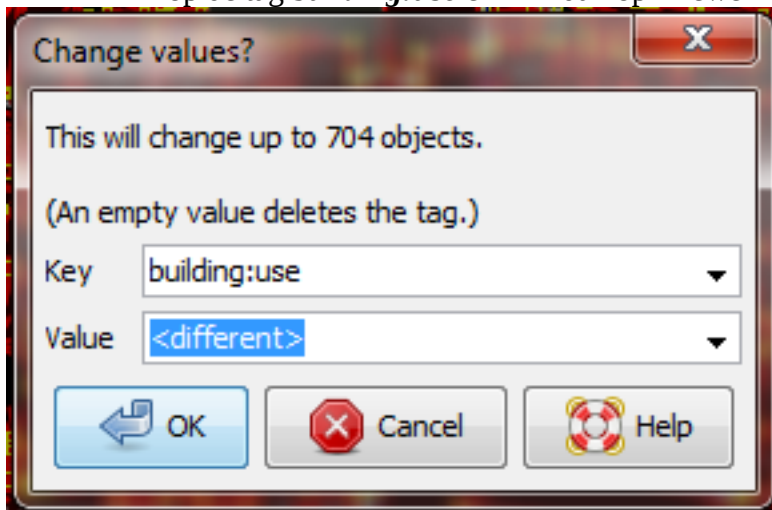
- Kijk in het venster Eigenschappen - wat we zien zijn alle tags voor elk geselecteerd object. Zij delen allemaal dezelfde sleutels, maar omdat elk object verschillende waarden

heeft zijn zij gemarkeerd als *<verschillend>*.



Key	Value
addr:housenumber	<different>
addr:street	<different>
amenity	<different>
building	<different>
building:condition	<different>
building:levels	<different>
building:material	<different>
building:structure	<different>
building:use	<different>
historic	<different>
name	<different>
religion	<different>
soft_storey	<different>
soft_storey:direction	<different>

- Klik op de tag **building:use** en klik dan op “Bewerken.”



Change values?

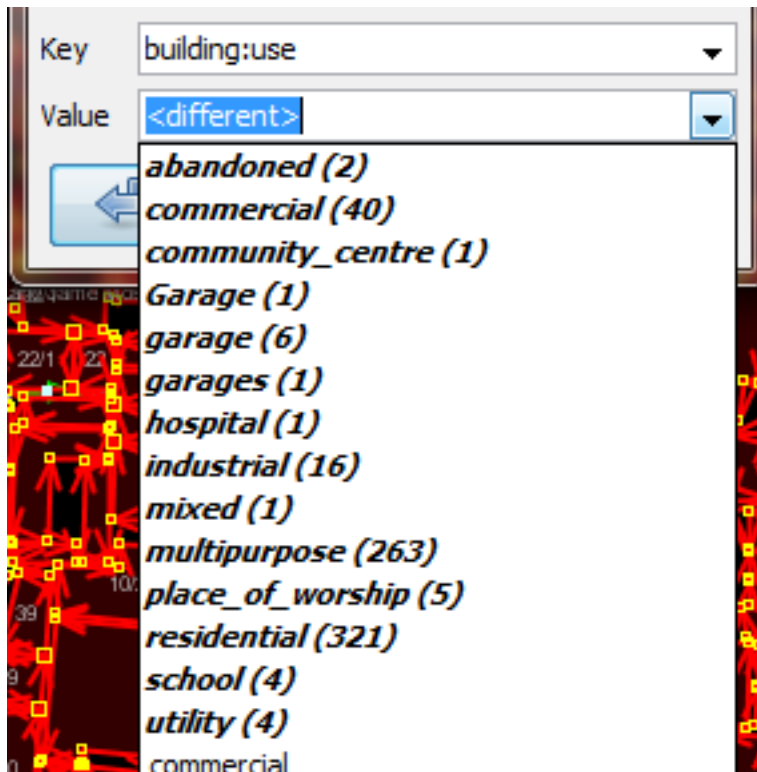
This will change up to 704 objects.
(An empty value deletes the tag.)

Key: building:use

Value: <different>

OK Cancel Help

- WEES HIER ZEER VOORZICHTIG! U wilt de waarde niet bewerken en op OK drukken omdat dat deze tag voor elk afzonderlijk object building zal wijzigen. **Dat zou heel slecht zijn.**
- Klik, in plaats daarvan, op het keuzeveld naast Waarde.



- Merk op dat alle items, in vet weergegeven, een nummer tussen haakjes naast zich hebben staan. Dat zijn het aantal geselecteerde objecten die deze tagwaarde hebben.

We kunnen die vergelijken met de tags van OpenStreetMap die in ons gegevensmodel in kaart zijn gebracht, en zoeken naar fouten. Deze tag geeft bijvoorbeeld het gebruik van het gebouw weer. Eerder in het project Open Cities Dhaka (waarvan deze gegevens afkomstig zijn) was er nog geen zekerheid of een gebouw met gemengd gebruik moest worden getagd als

building:use=multipurpose of ***building:use=mixed***. Omdat de eerste tag eerder in andere landen werd gebruikt, werd die geselecteerd. We zien hier echter dat één van de gebouwen is getagd als ***mixed***. We dienen dit te corrigeren. (Andere duidelijke fouten zijn de drie verschillende termen voor ***garage***, maar die zullen we hier niet corrigeren.)

- We kunnen het object, dat de tag ***building:use=mixed*** heeft, hier niet wijzigen, omdat we honderden objecten hebben geselecteerd. Dus, om de fout te kunnen corrigeren, moeten we hem eerst vinden. Hoe? U raadt het al - met het gereedschap Zoeken.
- Klik op “Annuleren” om dit dialoogvenster te verlaten. **Onthoud: klikken op OK kan gevaarlijk zijn.**
- Open Zoeken opnieuw en voer de volgende query in: “*building:use*=*mixed*”
- Onthoud dat de aanhalingstekens noodzakelijk zijn omdat een dubbele punt (:) ook een betekenis heeft bij het zoeken. Dit zou er voor moeten zorgen dat slechts het ene gebouw dat deze tag heeft wordt geselecteerd. Die kan nu worden gecorrigeerd naar de waarde ***multipurpose***.

Onthoud dat, als u de handleiding volgt, u N I E T moet proberen uw wijzigingen op te slaan in OpenStreetMap. Deze oefeningen zijn alleen ter demonstratie.

Opnieuw onderzoeken

Bij het beheren van een project zoals een gedetailleerd onderzoek van een gebouw, zou er een aanvullende methode voor kwaliteitsbeheer moeten zijn, zowel voor het verbeteren van het werk als voor het rapporteren van de nauwkeurigheid aan het einde van een project.

Als er veel teams zijn die in kaart brengen die samenwerken bij het onderzoeken van een gebied, komt het vaak voor dat één of meer teams niet naar behoren presteren. Zelfs de teams die efficiënt en nauwkeurig werken maken fouten. Veronderstel dat teams elk 100 gebouwen per dag in kaart brengen - het is niet onwaarschijnlijk dat een klein percentage van de attributen die zij verzamelen onjuist zouden kunnen zijn.

Dus, een goed project zal een proces bevatten voor het opnieuw onderzoeken van een gedeelte van het reeds uitgevoerde werk, repareren van fouten, bepalen welke teams naar behoren presteren, en een inschatting maken van het percentage fouten voor een eindrapport.

Natuurlijk heeft het geen zin in het opnieuw onderzoeken van elk gebouw in een doelgebied, maar 5-10% van de gebouwen zouden opnieuw onderzocht moeten worden. De na te kijken gebieden zouden moeten worden gekozen uit verschillende gebieden om te kunnen vergelijken tussen onderzoeksteams. Onderzoeksteams kunnen elkaars werk opnieuw onderzoeken, of mogelijk kunnen meer ervaren managers het nakijken op zich nemen. Het komt in de praktijk vaak voor dat managers één dag per week spenderen aan het opnieuw onderzoeken van delen van het doelgebied.

Fouten corrigeren

Wat zou er moeten worden gedaan als er fouten worden gevonden?

Als er een klein aantal fouten is (minder dan 5% van de gebouwen), zouden de problemen moeten worden voorgelegd aan het originele team dat ze in kaart heeft gebracht, zodat zij ervan weten en dezelfde fouten niet opnieuw maken. De gegevens zouden moeten worden gecorrigeerd in OpenStreetMap en de resultaten van het heronderzoek zouden moeten worden vastgelegd.

Als er meer fouten zijn, zouden andere acties kunnen worden ondernomen. Het onderzoeksteam zal op een nette manier dienen te worden aangesproken, en de gebieden die zij in kaart hebben gebracht zouden misschien wel in zijn geheel opnieuw moeten worden onderzocht, afhankelijk van hoe onbetrouwbaar de gegevens blijken te zijn. Grotere onnauwkeurigheid dan 10% is zeer waarschijnlijk een onacceptabele score.

Rapporteren over nauwkeurigheid

Het tweede doel van opnieuw onderzoeken is om te kunnen rapporteren over de nauwkeurigheid van de gegevens als het project wordt afgesloten. Gebruikers van de gegevens willen uw meetwijzen en methoden weten van het beoordelen van de kwaliteit van de gegevens.

Door dit proces op te nemen als deel van uw methode voor opnieuw onderzoeken, zult u in staat zijn helder uit te kunnen leggen hoe u de kwaliteit van de gegevens hebt beoordeeld, en harde getallen kunnen overleggen die het vermoedelijke foutpercentage weergeven die zijn opgenomen

in uw onderzoeksgegevens.

Laten we , als voorbeeld, veronderstellen dat we een project beheren dat 1000 gebouwen in kaart brengt. We besluiten om daarvan 10%, ofwel 100 gebouwen, in kaart te brengen, willekeurig geselecteerd uit het doelgebied. We gaan van start en stellen vast dat van de 100 gebouwen die we opnieuw hebben onderzocht, zes een hoog niveau van onnauwkeurigheid hebben. Laten we aannemen dat we onnauwkeurigheid definiëren als meer dan één attribuut niet juist. Dus zes procent in het opnieuw uitgevoerde onderzoek is fout - we kunnen deze fouten repareren, maar we moeten nog steeds extrapoleren dat ongeveer zes procent van alle 1000 gebouwen waarschijnlijk niet nauwkeurig zijn. Dat zou moeten worden gerapporteerd als de vermoedelijke foutmarge aan het einde van het project.

Opnieuw onderzoeken zou moeten worden uitgevoerd over het gehele project. Veronderstel dat we hebben gewacht tot het einde van dit voorbeeld en 40 van de 100 gebouwen zouden verkeerd zijn! Dat zou het gehele project kunnen ruïneren. Het is beter om grote fouten vroeg af te vangen zodat zij kunnen worden gecorrigeerd.

SQL-query's

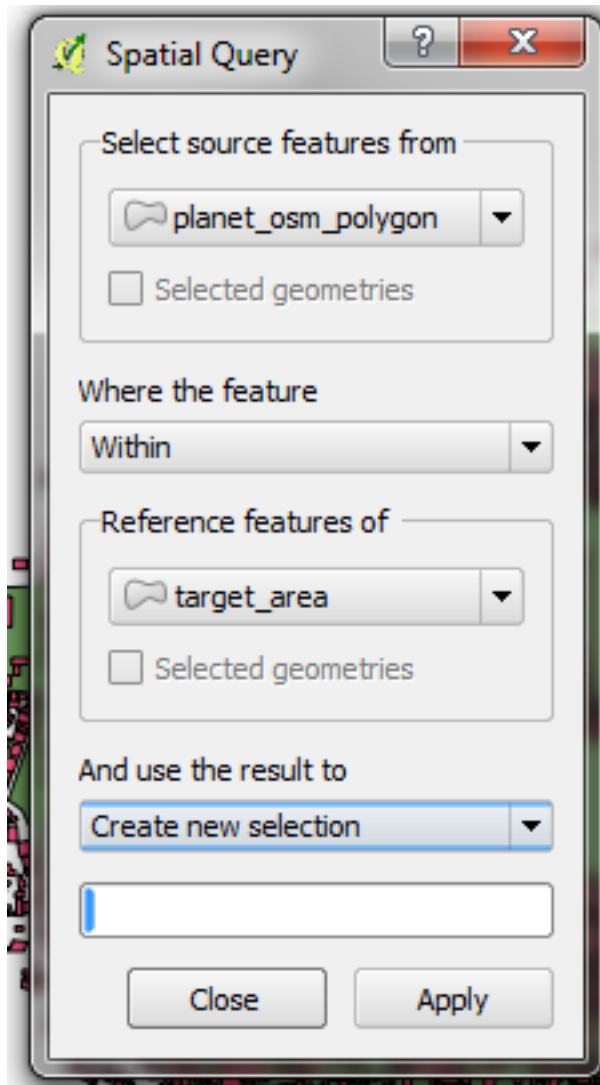
Het beste gereedschap voor analyse is waarschijnlijk het uitvoeren van SQL-query's in een GIS-systeem, zoals QGIS. Dat is soortgelijk aan het zoeken naar gegevens in JOSM, maar het biedt een meer krachtige analyse, hoewel het iets meer tijd kan vergen om in te stellen. Gebruiken van JOSM is een snelle, normale manier om te controleren of basisfouten, waar het uitvoeren van query's in QGIS beter is toegerust voor het zoeken naar ontbrekende gegevens of onjuiste attributen.

We gaan er hier van uit dat u al enigszins bekend bent met GIS, en focussen op het bouwen van query's die u kunnen helpen gegevens van OpenStreetMap na te kijken. Voor de oefeningen hieronder zullen we wederom gegevens gebruiken uit het project Open Cities Dhaka, die u kunt downloaden vanaf [dhaka_sql.zip](#). De gegevens van OpenStreetMap werden geëxporteerd met behulp van de HOT Export Tool ([export.hotosm.org](#)) en de grens van het doelgebied werd gedefinieerd aan het begin van het project.

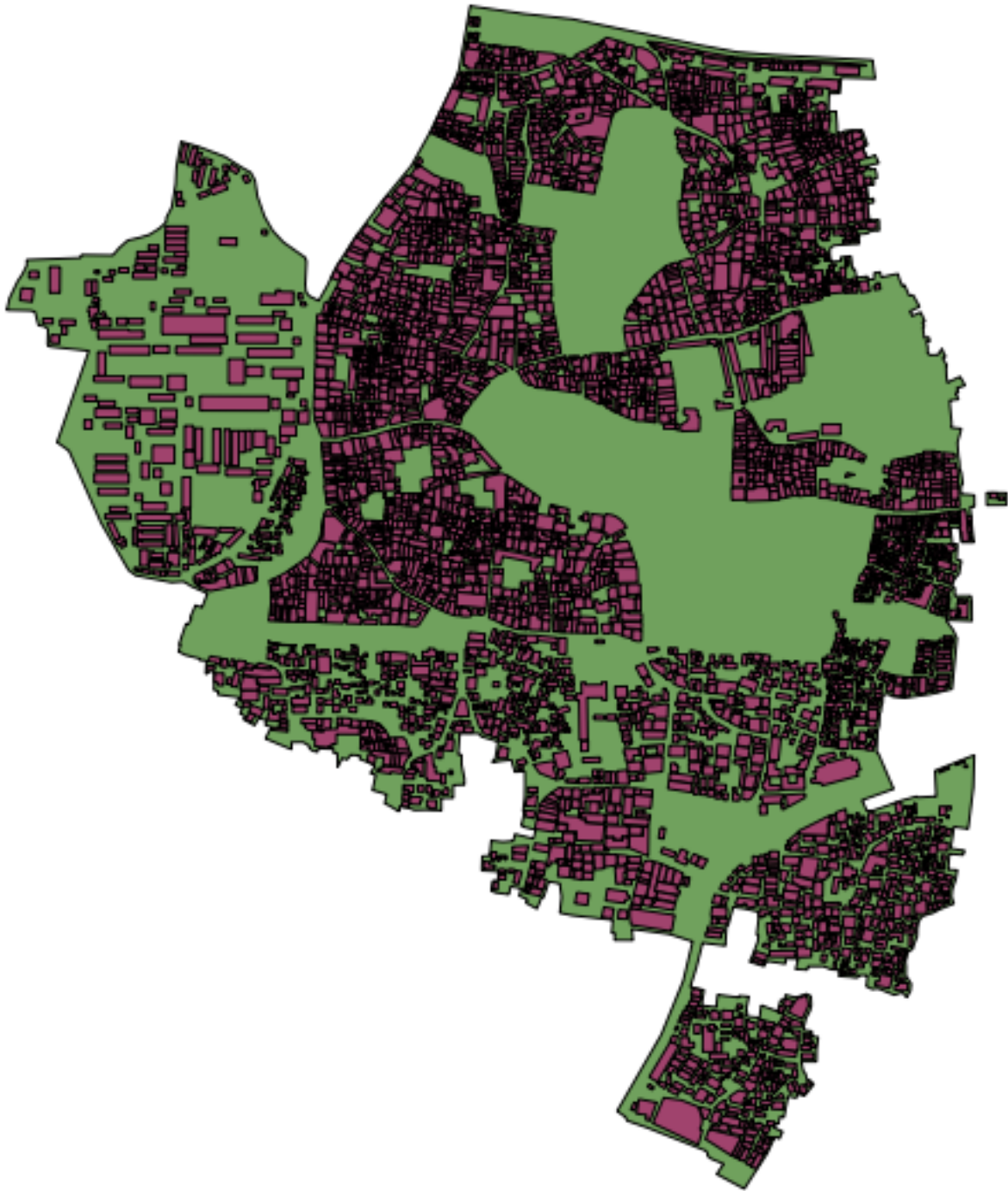
De gegevens voorbereiden

Pak de bestanden uit en laadt de twee shapefiles in QGIS. We zullen beginnen met het clippen van alleen de gebouwen binnen het projectgebied, om onze query's later eenvoudiger te kunnen maken.

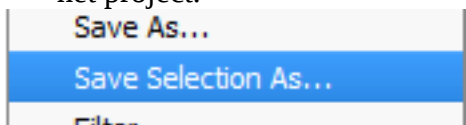
- Laten we eerst alleen de polygonen selecteren die liggen binnen het doelgebied. We zullen de plug-in Ruimtelijke Query gebruiken om dit te doen. Ga, als u die nog niet geïnstalleerd heeft, naar Plug-ins -> Plug-ins beheren en installeren om hem te zoeken en te installeren.
- Ga naar Vector -> Ruimtelijke Query -> Ruimtelijke Query.
- U zou instellingen moeten invullen om objecten te selecteren uit **planet_osm_polygon** die **within target_area** liggen.



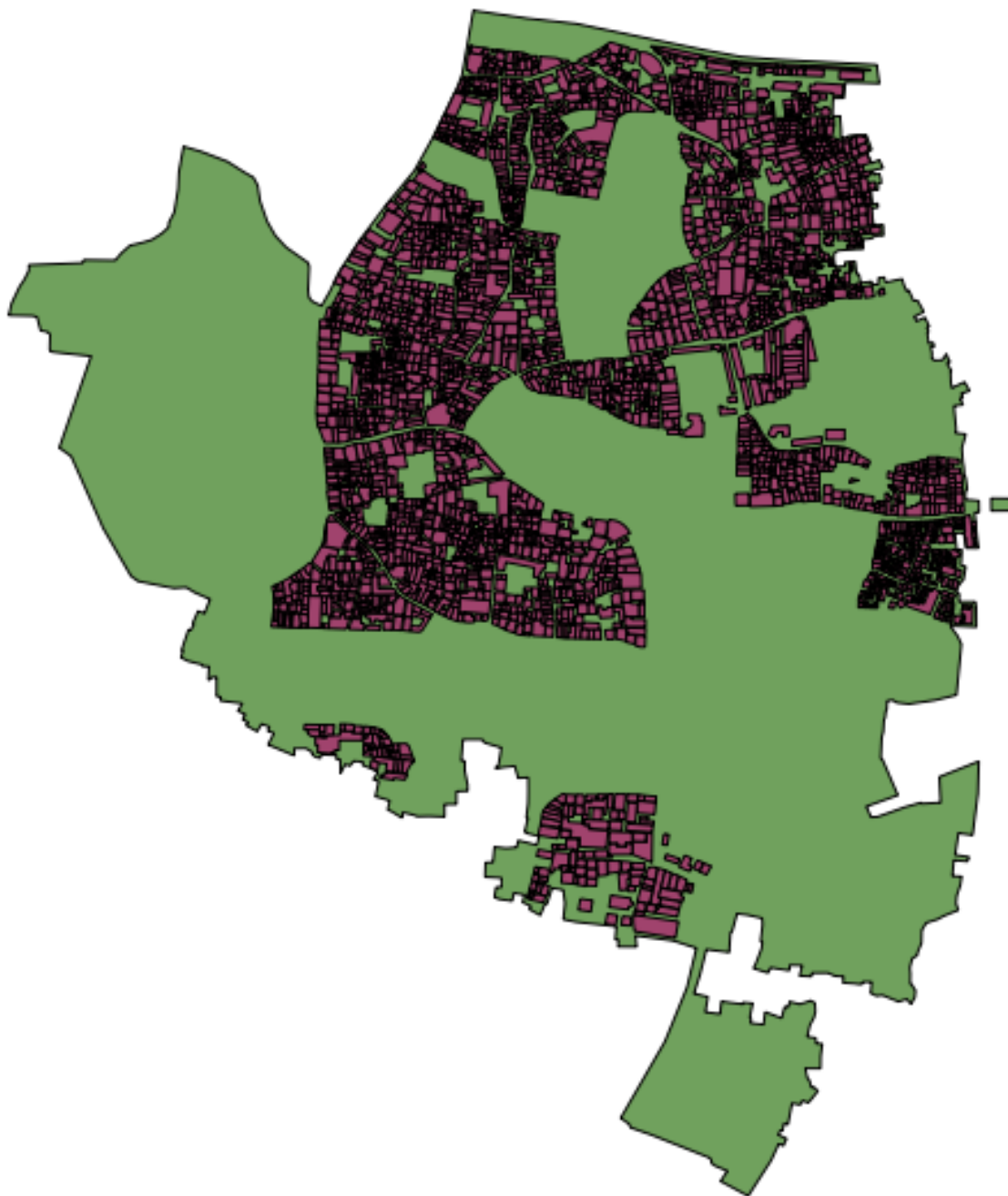
- Klik op Apply. Alleen polygonen die binnen het doelgebied liggen zullen worden geselecteerd.



- Klik met rechts op de laag en sla de selectie op als een nieuwe shapefile. Voeg die toe aan het project.



- Laten we vervolgens alleen de polygonen eruit filteren die gebouwen zijn en die werden verzameld als deel van dit project.
- Klik met rechts op **planet_osm_polygon** en klik op “Filter...”
- Voer de volgende query in:
“building” != NULL AND “source” = ‘Open Cities Dhaka Survey’
 - Klik op OK. Filteren van de gegevens met deze query zal alleen de polygonen weergeven die iets hebben vermeld in de kolom building. Het verwijdert ook gebouwen die niet de brontag hebben die is geassocieerd met dit project.
 - Sla deze gegevens op als een nieuw shapefile. We zullen dat bestand gebruiken voor onze SQL-query’s.



SQL-query's

We kunnen nu query's uitvoeren op de laag gebouwen om naar mogelijke fouten te zoeken. Laten we even denken aan enkele dingen die we zouden willen bevragen. Het gegevensmodel van dit project geeft attributen aan die zouden moeten worden verzameld voor elk gebouw - dat zijn:

- name
- building
- building:levels
- building:use
- building:vertical_irregularity
- building:soft_storey
- building:material
- building:structure
- start_date
- building:condition

Onthoud dat in de shapefile deze namen voor attributen zijn afgebroken, omdat benoemde kolommen zijn begrensd tot 10 tekens.

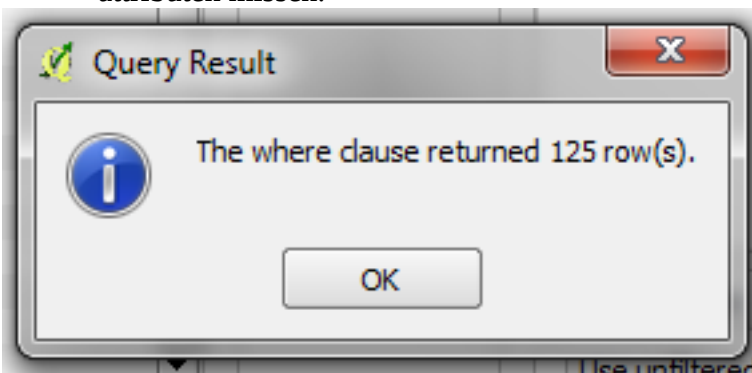
Dus, welk soort vragen willen we stellen? Wat zijn zeer waarschijnlijk fouten? Een veel voorkomende fout is dat een gebouw in kaart werd gebracht, maar niet alle attributen werden verzameld. We zullen dus een query willen uitvoeren die alle gebouwen weergeeft die geen volledige set attributen heeft. Natuurlijk is het voor bepaalde attributen, zoals name en start_date (construction year), prima dat zij leeg zijn, omdat niet elk gebouw een naam heeft en soms is het bouwjaar onbekend. Maar de andere attributen zouden altijd moeten zijn.

Laten we proberen daar een query voor te ontwikkelen:

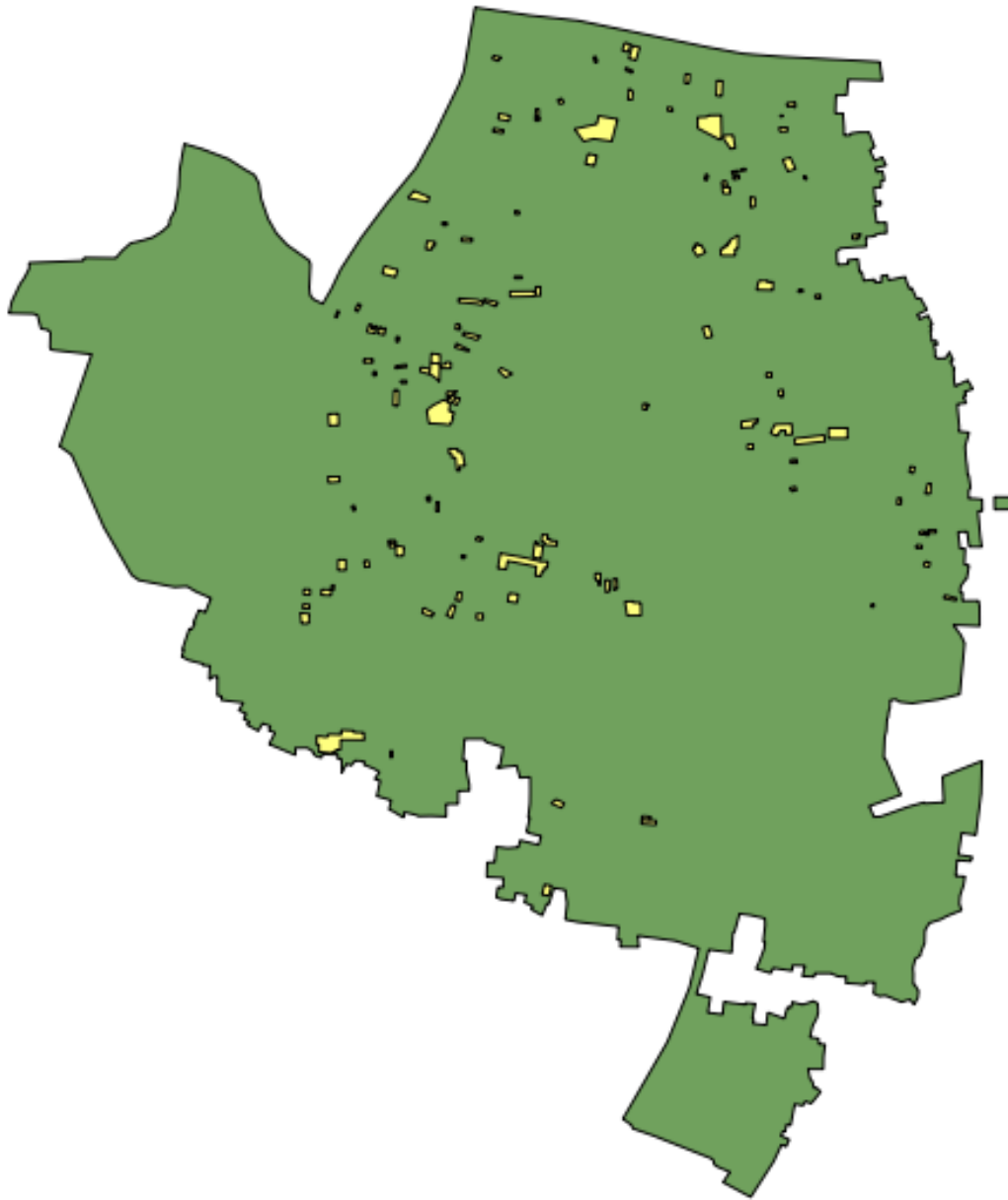
- Klik met rechts op de laag Gebouwen (de laag die we in het vorige gedeelte hebben gemaakt) en klik op “Filter...” Dit zal de Querybouwer openen. Hier kunnen we complex query's schrijven om alleen die gegevens terug te geven die we willen.
- U kunt uw eigen query bouwen door dubbel te klikken op de velden, operatoren, en waarden, of u kunt de query kopiëren die we hier hebben gebouwd:

“building_c” = NULL OR “building_s” = NULL OR “building_l” = NULL OR “building_m” = NULL OR “vertical_i” = NULL OR “soft_store” = NULL OR “building_u” = NULL

- Klik op “Test” en u zult zien dat deze query 125 objecten teruggeeft. Dat betekent dat van de ongeveer 3500 gebouwen die in kaart zijn gebracht, 125 daarvan één of meer van deze attributen missen.



- Klik op OK om alleen de gebouwen weer te geven die voldoen aan de voorwaarden in de query.



- Deze gebouwen kunnen dan nader worden bekeken om te identificeren welke attributen ontbreken, en of zij opnieuw dienen te worden onderzocht. Het is dan mogelijk om QGIS te gebruiken om een kaart te maken die het team dat het nieuwe onderzoek gaat uitvoeren kan gebruiken om de ontbrekende attributen voor de gebouwen te corrigeren.

Wat zijn andere query's die nuttig kunnen zijn? Wel, u wilt misschien ook controleren op attributen die niet zijn opgenomen in uw gegevensschema. We deden dit al in het gedeelte over het zoeken in JOSM. U kunt een query gebruiken om alle gebouwen te zoeken waarvan de attributen niet passen binnen uw gegevensmodel.

U kunt dit ook gebruiken voor het zoeken naar afwijkingen, die waarschijnlijk maar niet

noodzakelijkerwijze fouten zijn. Als we, bijvoorbeeld de Querybouwer openen, **building_1** selecteren, en klikken op “Alles” om alle mogelijke waarden voor attributen te laden, zien we dat de meeste gebouwen een getal hebben tussen één en 20 (Dit attribuut is building:levels, het aantal verdiepingen in de gebouwen). Maar er staat daar ook een 51 bij. Het is onwaarschijnlijk dat er een gebouw met 51 verdiepingen boven alles uitsteekt in dit gebied, dus we kunnen dat lokaliseren en er een notitie van maken om het te laten controleren bij degenen die het in kaart brachten.

Query's kunnen een effectieve manier zijn om te zoeken naar mogelijke fouten in de set met gegevens. Gecombineerd met andere mogelijkheden van QGIS, kan het worden gebruikt om kaarten te maken die kunnen worden gebruikt voor het nakijken van de gegevens in een gebied.

Samenvatting

In deze handleiding zijn we door een aantal effectieve methoden gegaan voor het onderhouden van de kwaliteit van de gegevens gedurende een project en hebben enkele hands-on oefeningen gedaan om het nakijken van de gegevens van OSM in de praktijk te brengen. Bij het organiseren van een project voor in kaart brengen, of zelfs bij het beoordelen van gegevens in een gebied voor persoonlijk gebruik, zouden deze methoden van nut kunnen zijn.

Was dit een goede handleiding? [Laat ons weten hoe we de handleidingen kunnen verbeteren!](#)

- learnosm@hotosm.org
- [@learnOSM](#)
- [Hosted on Github](#)



Official [HOT OSM](#) learning materials



Humanitarian
OpenStreetMap
Team